

Magyar Andrea

SZTE Neveléstudományi Doktori Iskola

Számítógépes adaptív tesztelés

A technológia fejlődésével a papír-ceruza tesztek (PP) használatát fokozatosan felváltja a technológia-alapú mérés-értékelés (Technology Based Assessment, TBA) (van der Linden, 2000). Ennek legelterjedtebb formája a számítógép-alapú tesztelés (Computer Based Testing, CBT), ahol a tanulók tesztelése számítógép előtt folyik, ami számos előnnyel jár mind a vizsgázók, mind a vizsgáztatók számára (Csapó, Molnár és R. Tóth, 2008). Többek között megbízhatóbbá válik a tesztelés, az eredmények azonnal hozzáférhetőek, a mérési idő lerövidül, és a vizsgáztatás egyéb körülményei, kontextuális információi is mérhetővé válnak (Csapó, Lőrincz és Molnár, megjelenés alatt). A számítógépes tesztelésre való áttérés történhet a PP tesztek digitalizálásával. Ebben az esetben csak a tesztet közvetítő médium változik meg, a feladatok sorrendje lineáris marad (Fix Item Testing, FIT) (Csapó, Molnár és R. Tóth, 2008). A számítógép adta lehetőségek igazi kihasználása az adaptív tesztek (Computerized Adaptive Testing, CAT) alkalmazásával valósulhat meg (Kárpáti, Molnár, Tóth és Főző, 2008). Adaptív tesztelés során minden vizsgázó a saját képességszintjéhez illeszkedő tesztet kap, ezáltal személyre szabottá válik a tesztelés (Way, 2006).

Az adaptív tesztelés ötlete már a századforduló első éveiben felvetődött, gyakorlati megvalósítása csak a számítógépek elterjedésével vált lehetővé (Linacre, 2000). Használata nemzetközi viszonylatban egyre inkább terjed, és egyre több cég használnál részben vagy egészében adaptív blokkokat. Elsősorban az Egyesült Államokban, de Európa mind több országában is használnak adaptív teszteket a tanulók vizsgáztatásra, képességmérésére (Meijer és Nering, 1999).

A CAT egyre nagyobb mérvű elterjedésének lehetősége teszi indokolttá a téma mélyebb vizsgálatát. A tanulmány áttekintést ad az adaptív tesztelés történeti előzményeiről, jelenlegi alkalmazásairól és az adaptív tesztelés folyamatáról, végül bemutatja a CAT főbb típusait és alkalmazási területeit.

Történeti előzmények

Az adaptív tesztelés eredete 1905-re vezethető vissza és Alfred Binet IQ-tesztjeinek kidolgozásához kötődik (van der Linden, 2000). Binet elsősorban egyéni diagnózissal foglalkozott, nem várhatta el, hogy mindenki ugyanazt a szintet teljesítse. Kidolgozott egy stratégiát, mellyel személyre szabottá tehetette a vizsgát: a gyerekek életkora alapján sorolta be az itemeket. A vizsgálatot a gyerek korának megfelelő kérdéssel kezdte, és a tesztelés során a vizsgázó képességeihez mérten addig kapott egyre nehezebb kérdése-

ket, amíg szintje bemérhetővé vált (*Linacre*, 2000). Ez a módszer azonban csak egyéni tesztelést tett lehetővé. A tesztelés nagy mintán való megvalósítása először a hadsereg keretein belül történt (Office of Naval Research; Navy Personnel Research and Development Center; Air Force Human Resources Laboratory; Army Research Institute) (*Wainer*, 2000). Hamar felismerték a CAT előnyeit, és széleskörű elméleti kutatásokat indítottak, melynek során számos, adaptív tesztelést megvalósító pszichometriai program került kifejlesztésre (ASVAB). A CAT igazán széles körű alkalmazása azonban a számítógépek elterjedésével vált lehetővé.

Frederic Lord volt az első az 1960-as évek végén, aki számítógépen valósította meg az item nehézség szintjének a vizsgázó képességszintjével való összehangolását (*Way*, 2006). A vizsga egy közepes nehézségi szintű itemmel kezdődött. Ha a vizsgázó helyesen válaszolt, kissé nehezebb kérdést kapott, ha viszont helytelen választ adott, akkor kissé könnyebb kérdéssel folytathatta. A vizsga addig folytatódott, míg a szintje beméréséhez elegendő számú itemre tudott válaszolni (*Weiss*, 2004).

1973-ban Weiss dolgozott ki egy Binet logikáján alapuló számítógépes tesztet, melyet 'stratified adaptive' vagy 'stradaptív' tesztnek nevezett. A teszt Binet-éhez hasonló itembankot használt, az itemeket szintenként csoportosítva, a különbség az itemek közvetítésében és a teszt befejezésében volt (*Weiss*, 1973).

Wainer és Kiely (1987) fejlesztette tovább az adaptív tesztek; itemek helyett úgynevezett tesztleteket használtak. A tesztlet rövid, néhány itemből álló kérdések csoportja. Ezek voltak az első nem item-alapú adaptív tesztek, használatuk később igen népszerűvé vált (*Wainer és Kiely*, 1987).

Az első adaptív teszteknél a klasszikus tesztelmélet módszereit használták az itemek kalibrációjára, a valószínűségi tesztelmélet (Item Response Theory, IRT) elterjedése óta e módszerek segítségével kalibrálják be az itemek nehézségi szintjét, mely jóval pontosabb kalibrálást tesz lehetővé (*Eggen*, 2004; *Al-A'ali*, 2007; *Weiss*, 2011).

Jelenlegi alkalmazások

Nemzetközi viszonylatban egyre elterjedtebb a CAT (*van der Linden*, 2000). A világ számos egyetemén foglalkoznak kutatócsoportok adaptív tesztekre irányuló kutatásokkal. Amerikában a minnesotai és a kansasi egyetemeken folynak jelentős kutatások. Európát tekintve Hollandiában a CITO és a Twente kutatócsoportja, Németországban az Aachen Egyetem RWTH kutatócsoportja, Angliában az NFER kutatócsoportja végez ilyen irányú vizsgálatokat. Az adaptív tesztekkel foglalkozó kutatók nemzetközi szervezete, az IACAT (International Association for Computerized Adaptive Testing) 2010-ben alakult, és célja az adaptív tesztelés pszichológia és oktatás terén történő fejlesztése és terjesztése. Az adaptív tesztek használata terén jelenleg az Amerikai Egyesült Államok jár az élen (*Weiss*, 2004). Az egyik legismertebb nagymintán alkalmazott teszt, a Graduate Management Admission Test (GMAT), az üzleti iskolák számára került kifejlesztésre. Kifejlesztője és működtetője az amerikai Graduate Management Admission Council, de ma már a világ több, mint 150 országában elérhető (*Frey és Seitz*, 2009; *Wainer*, 2000). Szintén amerikai a COMPASS tesztkészlet, melyet az American College Test (ACT) program során fejlesztettek ki, és a főiskolákra jelentkezők matematika és olvasáskészségét méri fel.

További amerikai adaptív felmérések a Measures for Academic Progress (MAP), a Virginia Standards of Learning Tests (SOL), valamint az Oregon Assessments of Knowledge and Skills (OAKS) (*Molnár*, 2010). Több amerikai tesztnek létezik adaptív verziója is: Graduate Record Examination, Computerized Placement Test (*Meijer és Nering*, 1999).

Európában kisebbek az online tesztelés hagyományai, de az utóbbi években itt is egyre több felmérésben alkalmaznak adaptív tesztek (*Molnár*, 2010). A Prog-

ram for International Assessment of Adult Competence (PIAAC) felmérése a felnőtt lakosság szövegértési és matematikai képességeit méri fel. A teszt 25 perces adaptív blokkokat tartalmaz, a számítógép a megoldás alapján adagolja a következő nehézségű blokkot. A TOEFL nyelvvizsga is adaptív rendszerű: összesen 4–4,5 óras számítógépes adaptív teszttel méri a vizsgázók nyelvtani ismeretét, szókincsét, hallás utáni szövegértését és íráskészségét (Molnár, 2010). A holland CITO számos, az oktatás területén alkalmazandó adaptív tesztet dolgozott ki. A tesztek között szerepelnek matematikai, számolási, szókincsfelmérési tesztek, 4 éves kortól egészen felnőttkorig.

Az OECD PISA a legjelentősebb rendszerszintű, az oktatásban használatos felmérés. A felmérés három fő területet (szövegértés, matematika, természettudomány) mér, melyek három évenként ciklikusan ismétlődnek. A mérések jelenleg papíralapon folynak, de az OECD kísérleti méréseket tervez bevezetni az adaptív tesztelés megvalósítására vonatkozóan (Frei és Seitz, 2009). A tervek szerint a 2015-ös mérést a Pearson cég készíti elő, és várhatóan adaptív blokkokat is fog tartalmazni.

Hazánkban is léteznek adaptív modulokat alkalmazó tesztek. Ilyen a Vienna Test System magyarul is elérhető értékelő eszköze, mely a pszichológiai vizsgálatok széles területén alkalmazható, úgymint a klinikai, sport-, repülés-, személyiség-, neuro- és közlekedés-pszichológia. Az oktatás területén nagymintás vizsgálatokban is alkalmazható adaptív mérőeszköz bevezetése jelenleg az SZTE OK által folyamatban van (R. Tóth, Molnár, Latour és Csapó, 2011).

Az adaptív tesztek szigorú algoritmus szerint működnek

(Linacre, 2000). A teszt kezdetén kiválasztásra kerül a kezdő item/stage amennyiben előzetes információk rendelkezésre állnak a vizsgázóról, akkor ezek felhasználásával, ha nem, akkor véletlenszerű kiválasztással az itembankból vagy annak részhalmazából. Rendszerint egy közepes nehézségű itemmel indul a tesztelés. Az adott item megoldása után egy újabb kiválasztása következik. Ha a vizsgázó jól válaszolt, egy nehezebbet kap, amennyiben hibázott, akkor könnyebbet (1. ábra). A program algoritmusa biztosítja, hogy minden soron következő item a vizsgázó képességeihez mért legyen.

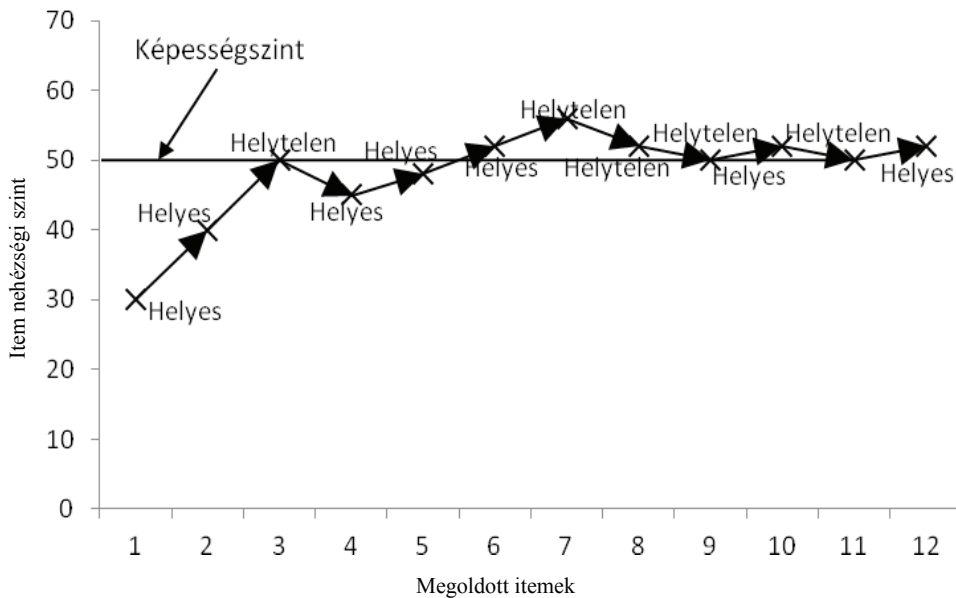
Az adaptív tesztelés folyamata

A megoldott itemeket a gép értékeli, és dönt arról, hogy szükséges-e új item kiválasztása, vagy a tesztelés véget ért. A tesztelés végén a vizsgázó azonnali visszajelzést kap elért eredményéről (Csapó, Molnár és R. Tóth, 2008; Eggen, 2004). Ez alapján a CAT individualizált teszt: a vizsgázók különböző itemekkel kezdenek és folytathatják a tesztet, és különböző lehet a megoldott itemek száma is. A CAT dinamikus, személyre szabott, a vizsgázó szintjéhez igazodik (Weiss, 2011).

A CAT előnyei és hátrányai

Az adaptív tesztek számos fontos előnnyel rendelkeznek a FIT tesztekhez képest. Linacre (2000), Wainer (2000), Eggen (2007), Tian (2007) és Cisar (2010) az adaptív tesztek következő előnyeit emelik ki:

- Biztonságosabbá válik a tesztelés. Mivel minden vizsgázó különböző kérdéseket kap, nincs lehetőség a kérdések előzetes „betanulására” (Wainer, 2000; Eggen, 2007; Tian, 2007).



1. ábra. Adaptív tesztelés (Linacre, 2000 alapján)

- Minden vizsgázó a saját sebességének megfelelően dolgozhat (a válaszadás gyorsasága további kiegészítő információ lehet) (Linacre, 2000; Tian, 2007).
- A vizsgázók saját szintjüknek megfelelő feladatokat kapnak. Így kiküszöbölhető, hogy a vizsgázók túl könnyű, illetve túl nehéz kérdéseket kapjanak, és ezáltal unalmassá váljon számukra a vizsga, valamint csökken a frusztráció (Wainer, 2000; Eggen, 2007).
- A tesztelési idő lerövidül, így kevésbé fárasztó a vizsgázók számára, és az értékelés is gyorsabbá, egyszerűbbé válik (Linacre, 2000; Tian, 2007).
- Az adaptív tesztelés individualizált jellege miatt a pontosabb képesség-felmérés lehetőségét teremti meg (Linacre, 2000).

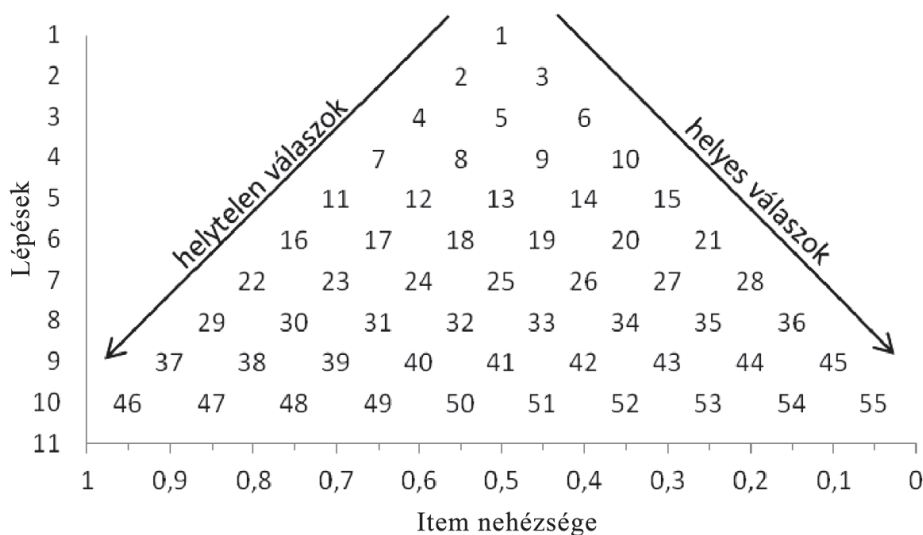
Azonban a CAT hátrányokkal és korlátokkal is rendelkezik, melyek közül a legfontosab-
bak (Linacre, 2000; Wainer, 1987; Wainer, 2000; Meijer és Nering, 1999; Cisar, 2010):

- Az adaptív teszteknek jóval nagyobb az előállítási költségük, mint a fix teszteknek (Linacre, 2000; Wainer, 2000; Meijer és Nering, 2000).
- Nincs lehetőség a tesztek visszamenőleges áttekintésére (Linacre, 2000; Wainer, 2000).
- A lineáris teszteknel a vizsgázónak lehetősége van kérdéseket kihagyni, és utólag visszatérni rájuk. A legtöbb adaptív teszt nem engedi ezt a lehetőséget (Wainer, 2000; Cisar, 2010).
- Az itemek nagyon precíz kalibrálást igényelnek (Linacre, 2000; Cisar, 2010).
- Az item elhelyezkedése befolyásolhatja a megoldást: ugyanaz az item a tesztben elfoglalt helyétől függően lehet könnyebben, illetve nehezebben megoldható (Wainer, 1987).
- Az előző item véletlenül információt szolgáltat a következő item számára (kereszt-információ) (Wainer, 1987).
- Az adaptív tesztek nem alkalmasak minden tantárgy és képesség mérésére (Linacre, 2000; Cisar, 2010).

A CAT típusai

A felmérések céljától függően többféle adaptív tesztípus került kifejlesztésre. Az item-alapú vagy teljes adaptív teszt esetében a következő item csak akkor választódik ki, mikor az előzőt megválaszták. Ilyenek voltak a legkorábbi, például a Lord által készített, úgynevezett rugalmas szintű (flexilevel) tesztek (Lord, 1971; Pyper, 2010) és az elágazó, sorozatos vagy piramis-tesztek ('branched', 'sequential', 'pyramidal') (Larkin és Weiss, 1974).

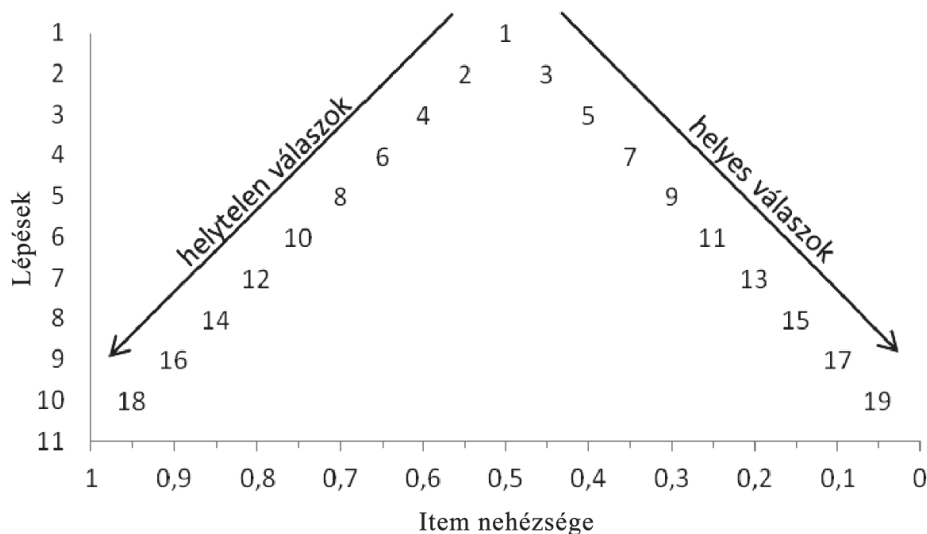
A 2. ábra egy 10 lépcsős, állandó lépésközű piramis-teszt sematikus ábrázolását mutatja (Weiss, 1974 alapján).



2. ábra. Piramis-teszt struktúrája (Weiss, 1974 alapján)

Az első item a piramis csúcsán helyezkedik el, és közepes nehézségi szintű. A második item az első megoldásától függően lehet nehezebb vagy könnyebb, vagyis kétféle. Ennek megfelelően a harmadik szinten háromféle nehézségű item helyezkedik el, és így tovább, míg a teszt véget nem ér, így minden szinten a szint sorszámával megegyező mennyiségű itemre van szükség. Egy tízlépcsős piramis 55 itemet igényel. Az állandó lépésközű ('constant step size') piramissnál az itemek nehézségi foka mindig egyenlő mértékben változik, míg a csökkenő lépésközű ('decreasing step size') piramissnál az elején nagyobb különbségek vannak az itemek nehézségi fokában, a végén pedig egyre kisebbek. Ezzel a módszerrel jelentős időmegtakarítás érhető el (Weiss, 1974).

A rugalmas szintű tesztek jelentik a minimalista adaptív tesztek alapját (3. ábra). Ezek abban különböznek a piramis-tesztektől, hogy minden item különböző nehézségi szintű, egy-egy item megválaszolása után a nehézségi szintben legközelebb álló, meg nem választott item következik (Betz és Weiss, 1975). Így a piramis-tesztekhez képest jóval kevesebb számú item (10 lépcsős tesztnél 19) elegendő. Az értékelés is egyszerűbb, mivel tulajdonképpen minden vizsgázó, aki ugyanazt a pontszámot érte el, ugyanazt a tesztet írta. E tesztek alkalmazásának hátránya, hogy az alacsony itemszám miatt a mérés precizitása erősen gyengülhet (Pyper és Lilley, 2010).



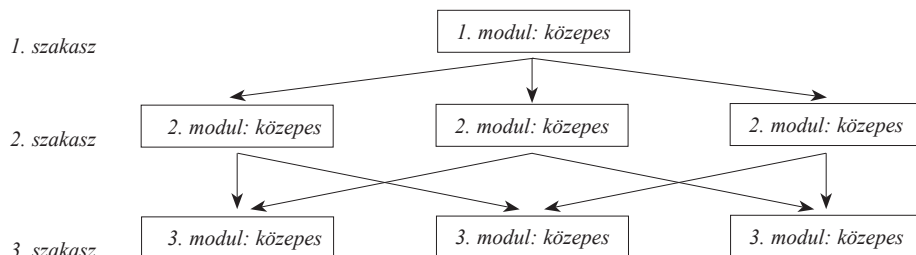
3. ábra. Rugalmas szintű teszt (Betz és Weiss, 1975 alapján)

Az önadaptív teszt (Self Adaptive Test, SAT) annyiban különbözik a fent említett flexilevel vagy piramis-tesztekkel, hogy a vizsgázó minden egyes lépés után maga dönthet, hogy a következő lépésben a könnyebb vagy nehezebb itemet kéri (Vispoel, 1998). Nagy előnye ennek a tesztelésnek, hogy jelentősen csökkenti a vizsgadrukkot (Wise, Plake, Johnson és Roos, 1992; Roos, Wise és Plake, 1997). Pitkin és Vispoel (2001) viszont arra a következtetésre jutottak, hogy habár a SAT valóban csökkenti a vizsgadrukkot, valamint növeli az elért pontszámokat, viszont kevésbé megbízható. Mivel a vizsgázó maga döntheti el, hogy a következő lépésben milyen nehézségű itemet kér, gyakran nem a képességszintjéhez mérten, hanem izgalmi állapotához viszonyítva választ. Másik fontos hátránya, hogy jelentősen megnöveli a tesztelés időtartamát.

A rétegzett-adaptív ('stratified-adaptive') vagy stradaptív tesztekkel az itemek nehézségi szintjüknek megfelelően rétegekbe ('strata') vannak csoportosítva. A tesztelés az előző típusoktól eltérően nem közepes szinten, hanem a vizsgázótól előzetesen bekért információk alapján a vizsgázó képességszintjén kezdődik. Egy-egy item megoldása után a vizsgázó a legközelebbi réteg meg nem oldott feladatát kapja. Ezeknek a teszteknek egyedi tulajdonsága, hogy az itemek sorrendje előre kiszámíthatatlan, sokkal több variáció lehetséges, mint a piramis- vagy a flexilevel tesztekkel, és ezáltal precízebb mérést tesztek lehetővé (Weiss, 1973).

A nem item-alapú tesztek közül a legegyszerűbb szerkezetűek az úgynevezett kétszakaszos ('two-stage') tesztek. Ezeknél először minden vizsgázó ugyanazt a kezdő tesztet ('routing test') kapja, mely vegyesen tartalmaz könnyebb és nehezebb itemeket. Ezen a teszten elért eredménye alapján kapja a vizsgázó a második részben a saját szintjéhez illeszkedő felmérő tesztet ('measurement test') (Weiss, 1974).

A többszakaszos tesztekkel (Multi-Stage Test, MST) ez több szakaszban ismétlődik (4. ábra). Annak ellenére, hogy az item-alapú adaptív teszt esetében feleannyi item elegendő, mint a szakaszos tesztekkel, az utóbbi években mégis a többszakaszos tesztek váltak népszerűbbé (Eggen, 2007). Számos nagymintás teszt alkalmaz többszakaszos tesztet, például a NAEP (National Assessment of Educational Progress) vagy a TOEFL (Test of English as a Foreign Language) (Keng, 2008).



4. ábra: Többszakaszos teszt struktúrája (Keng, 2008 alapján)

További nem item-alapú tesztek Wainer és Kiely (1987) tesztjei, melyeknél itemek helyett tesztleteket javasoltak a kiválasztás alapjául. A tesztlet egy rövid, néhány itemből álló összefüggő teszt, mint például egy rövid szövegre vagy egy térképre, táblázatra vagy grafikonra vonatkozó tesztkérdések csoportja (Keng, 2000). Wainer, Bradlow, és Du (2000) több előnyét említi a tesztletek alkalmazásának: egyrészt csökkenti a különálló, kis itemek feletti aggályokat, másrészt időhatékony, mivel egy szöveghez vagy táblázathoz számos kérdés tartozhat. A tesztletek egyik alkalmazása a CAST (Computer Adaptive Sequential Testing), ahol a vizsga egy közepes nehézségű tesztlettel kezdődik, és ezt követően a vizsgázó választától függően kap könnyebb, nehezebb vagy az előzővel megegyező nehézségű tesztletet (Luecht, 2000). Az MSF (Multiple Form Structure) szintén tesztleteket alkalmazó CAT (Armstrong, 2004).

Összefoglalás

A technológia elterjedése nemcsak mindennapi életünket, hanem az oktatás folyamatát is jelentős mértékben befolyásolja. A mérés-értékelés területén is jelentős változások történtek, a hagyományos tesztelést egyre inkább felváltja a számítógép-alapú mérés, melynek ma ismert legfejlettebb formája a számítógép-alapú adaptív tesztelés. Nemzetközi viszonylatban egyre több egyetemen foglalkoznak kutatócsoportok adaptív tesztek fejlesztésével, és több nagymintás tesztnek dolgozták ki az adaptív verzióját.

Adaptív tesztelés során a vizsgázók a saját képességszintjüknek megfelelő feladatokat kapnak, és az értékelés is az egyéni teljesítmények alapján történik. Így számos előnnyel rendelkeznek a fix tesztekkel szemben, a tesztelés a vizsgázó képességszintjéhez alakítható, lerövidül a tesztelési idő és az értékelés is gyorsabbá, egyszerűbbé válik.

Azonban vannak a CAT bevezetésének korlátai is. Akár papír-ceruza tesztről történik az átállás, akár számítógépes fix tesztről, számos tényezőt kell figyelembe venni, hogy a kétféle teszt összehasonlíthatóságát biztosítani lehessen. Ilyenek a médiahatáson kívül az itemek különböző elhelyezkedése során fellépő nehezebb vagy könnyebb megoldhatóság, a kereszt-információ problémája, vagy az eltérő pontozási lehetőségek.

A nemzetközi kutatások során az adaptív tesztek számos típusát fejlesztették ki. Az item-alapú tesztek közül a legkorábbiak a rugalmas szintű, úgynevezett flexilevel tesztek és a piramis-tesztek voltak. Ezeknél a tesztelés során a következő item az előző itemre adott válasz függvényében került kiválasztásra. További item-alapú teszt az önadaptív teszt, melynek fő előnye, hogy jelentősen csökkenti a vizsgadrukkot, ezáltal növelve az elért teljesítményt. Papír és ceruza tesztektről való átállás során leggyakrabban a nem item-alapú tesztek vezetnek be, ezek közül is a kétszakaszos, többszakaszos és a tesztlet-alapú tesztek. A kétszakaszos teszteknél először minden vizsgázó ugyanazt a kezdő

tesztet kapja, majd a megoldás alapján kap a második részben egy képességéhez igazodó differenciált tesztet. A többszakaszos teszteknel ez több szakaszban ismétlődik. A tesztletek rövid, egy-egy táblázatra, grafikonra vagy szövegre vonatkozó tesztkérdések csoportjai. Alkalmazásuk előnye, hogy az item-alapú tesztek számos hátránya kiküszöbölhető velük, úgymint az item helyzetéből adódó eltérő nehézség vagy a keresztfinformáció.

A számítógépes adaptív tesztek új lehetőséget teremtenek a mérés-értékelés terén. Szélesebb körben való elterjedésük mind a pedagógia, mind a pszichológia területén számíthat olyan lehetőséget kínálnak, melyek illeszkednek a technológia adta lehetőségekhez és napjaink elvárásaihoz.

Irodalomjegyzék

- Al-A'ali, M. (2007): Implementation of an improved adaptive testing theory. *Educational Technology & Society*, 10. 4. sz. 80–94.
- Armstrong, R. D., Jones, D. H., Koppel, N. B. és Pashley, P. J. (2004): Computerized adaptive testing with multiple-form structures. *Applied Psychological Measurement*, 28. sz. 147–164.
- Betz, N. E. és Weiss, D. J. (1975): *Empirical and simulation studies of flexilevel ability testing*. (Research Report 75-3). University of Minnesota, Department of Psychology, Psychometric Methods Program, Minneapolis.
- Čisar, S. M., Radosav, D., Markoski, B., Pinter, R. és Čisar, P. (2010): Computer adaptive testing of student knowledge. *Acta Polytechnica Hungarica*, 7. 4. sz. 139–152.
- Csapó Benő, Molnár Gyöngyvér és R. Tóth Krisztina (2008): A papír alapú tesztetől a számítógépes adaptív tesztelésig: a pedagógiai mérés-értékelés technikájának fejlődési tendenciái. *Iskolakultúra*, 18. 3–4. sz. 3–16.
- Csapó Benő, Lőrincz András és Molnár Gyöngyvér (megjelenés alatt): Innovative assessment technologies in educational games designed for young students. In: Ifenthaler, D., Eseryel, D. és Ge, X.: *Assessment in game-based learning: foundations, innovations, and perspectives*. Springer, New York.
- Eggen, T. J. H. M. (2004): *Contributions to the theory and practice of computerized adaptive testing*. Citogroep Arnhem, Netherlands.
- Eggen, T. J. H. M. (2007): Choices in CAT models in the context of educational testing. In: Weiss, D. J. (szerk.): *Proceedings of the 2007 GMAC Conference on Computerized Adaptive Testing*.
- Frey, A. és Seitz, N. N. (2009): Multidimensional adaptive testing in educational and psychological measurement: Current state and future challenges. *Studies in Educational Evaluation*, 35. 2–3. sz. 89–94.
- Kárpáti Andrea, Molnár Gyöngyvér, Tóth Péter és Főző Attila (2008, szerk.): *A 21. század iskolája*. Nemzeti Tankönyvkiadó. Budapest.
- Keng, L., Tusng-Han, H., Tzu-An, C. és Dodd, B. (2000): *A comparison of item and testlet selection procedures in computerized adaptive testing*. New York, N.Y.
- Keng, L. (2008): *A Comparison of the performance of testlet-based computer adaptive tests and multistage tests*. The University of Texas, Austin.
- Larkin, K. C. és Weiss, D. J. (1974): *An empirical investigation of computer-administered pyramidal ability testing* (Research Report 74-3). University of Minnesota, Department of Psychology, Psychometric Methods Program, Minneapolis.
- Linacre, J. M. Ph.D. (2000): *Computer-adaptive testing: A methodology whose time has come*. MESA Psychometric Laboratory, University of Chicago.
- Lord, F. M. (1971): The self-scoring flexilevel test. *Journal of Educational Measurement*, 8. 3. sz. 147–151.
- Luecht, R. M. (2000): Implementing the computer-adaptive sequential testing (CAST) Framework to mass produce high quality computer-adaptive and mastery tests. Annual Meeting of the National Council on Measurement in Education. New Orleans, LA.
- Meijer, R. R. és Nering, M. R. (1999): Computerized adaptive testing: Overview and introduction. *Applied Psychological Measurement*, 23. sz. 187–194.
- Molnár Gyöngyvér (2010): Technológia-alapú mérés-értékelés hazai és nemzetközi implementációi. *Iskolakultúra*, 20. 7–8. sz. 22–34.
- Pitkin, A. K. és Vispoel, W. P. (2001): Differences between self-adapted and computerized adaptive tests: A meta-analysis. *Journal of Educational Measurement*, 38. 3. sz. 235–247.
- Pyper, A. és Lilley, M. (2010): *A comparison between the flexilevel and conventional approaches to objective testing*. CAA Konferencia, University of Hertfordshire.
- Roos, L. L., Wise, S. L. és Plake, B. S. (1997): The role of item feedback in self-adapted testing.

- Educational and Psychological Measurement*, **57**. 1. sz. 85–98.
- R. Tóth Krisztina, Molnár Gyöngyvér, Latour, T. és Csapó Benő (2011): Az online tesztelés lehetőségei és a TAO platform alkalmazása. *Új Pedagógiai Szemle*, **61**. 1–2–3–4–5. sz. 8–22.
- Tian, J., Miao, D. és Zhu Xia, G. J. (2007): An Introduction to the computerized adaptive testing. *US-China Education Review*, **4**. 1. sz.
- van der Linden, W. J. és Glas, C. A. W. (2000, szerk.): *Computerized adaptive testing: Theory and practice*. Kluwer, Boston.
- Vispoel, W. P. (1998): Reviewing and changing answers on computer-adaptive and self-adaptive vocabulary tests. *Journal of Educational Measurement*, **35**. 4. sz. 328–345.
- Wainer, H. és Kiely, G. (1987): Item clusters and computerized adaptive testing: A case for testlets. *Journal of Educational Measurement*, **24**. sz. 185–201.
- Wainer, H. (2000): *Computerized adaptive testing: A primer*. 2. kiadás. Erlbaum, Hillsdale, NJ.
- Wainer, H. (2000): CATs: Whither and whence. *Psicológica*, **21**. sz. 121–133.
- Wainer, H., Bradlow, E. T. és Du, Z. (2000): Testlet response theory: An analog for the 3PL model useful in testlet-based adaptive testing. In: van der Linden, W. J. és Glas, C. A. W.: *Computerized adaptive testing: Theory and practice*. Kluwer, Dordrecht, Netherlands. 245–269.
- Way, W. D., Davis, L. L. és Fitzpatrick, S. (2006): *Practical questions in introducing computerized adaptive testing for K-12 assessments*. Pearson.
- Weiss, D. J. (1973): *The stratified adaptive computerized ability test* (Research Report 73-3). Psychometric Methods Program, Computerized Adaptive Testing Laboratory, University of Minnesota, Department of Psychology, Minneapolis.
- Weiss, D. J. (1974): *Strategies of ability measurement*. (Research Report 74-x). Psychometric Research Program, Department of Psychology, University of Minnesota, Minneapolis.
- Weiss, D. J. (2004): Computerized adaptive testing for effective and efficient measurement in counseling and education. *Measurement and Evaluation in Counseling and Development*, **2**. sz. 70–84.
- Weiss, D. J. (2011): Better data from better measurements using computerized adaptive testing. *Journal of Methods and Measurement in the Social Sciences*, **2**. 1. sz. 1–27.
- Wise, S. L., Plake, S. S., Johnson, P. L. és Roos, S. L. (1992). A comparison of self-adapted and computerized adaptive achievement tests. *Journal of Educational Measurement*, **29**. sz. 329–339.